RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/698,213A

Input Set : **A:\SeqList.txt**
Output Set: **N:\CRF3\05212001\I698213A.raw**

```
 2 <110> APPLICANT: McIninch, James
 4 <120> TITLE OF INVENTION: COMPUTATIONAL NUCLEIC ACID CODING AND FEATURE ANALYSIS
 6 <130> FILE REFERENCE: 04983.0220.00US00
 8 <140> CURRENT APPLICATION NUMBER: 09/698,213A
10 <141> CURRENT FILING DATE: 2000-10-30
12 <160> NUMBER OF SEQ ID NOS: 13
14 <170> SOFTWARE: PatentIn version 3.0
16 <210> SEQ ID NO: 1
17 <211> LENGTH: 2165
18 <212> TYPE: DNA
19 <213> ORGANISM: Arabidopsis thaliana
21 <220> FEATURE:
22 <221> NAME/KEY: unsure
23 <222> LOCATION: (1)...(2165)
24 <223> OTHER INFORMATION: Unsure at all n locations
26 <220> FEATURE:
27 <223> OTHER INFORMATION: Ecotype Landsberg, genomic DNA
29 <400> SEQUENCE: 1
30 tactcaaaaa tatattccat gcttaattag gccggattcg cggtgacgat gcaccaagag      60
32 cggtttttcc gagcattgta ggccgtcctc gccacaccgg tgtgatggtt gggatgggac     120
34 aaaaggatgc ttatgttgga gacgaggctc aatcaaaacg tggtatcttg actctgaagt     180
36 acccaattga gcatggaatt gttaataatt gggatgacat ggagaagatt tggcatcaca     240
38 ctttctacaa tgagcttcgt gttgccctg aagaacatcc ggttctcttg accgaagctc      300
40 ctctcaatcc gaaagctaac cgtgagaaga tgactcagat catgtttgag acattcaata     360
42 ctcctgctat gtatgttgcc attcaagctg ttctctcact ctatgccagt ggccgtacta     420
44 ctggtcagta cattactaca ttcttttat accgtttggt tgaaataaaa ttcggtttgg      480
46 ttcgattcga gtttgctctc attattttta ttttgttggt taggtattgt tttggactcc     540
48 ggagatggtg tgagccacac ggtaccaatc tacgagggtt atgcacttcc acacgcaatc     600
50 ctgcgtcttg atcttgcagg tcgtgaccta accgaccacc ttatgaaaat cctgacagag     660
52 cgtggttact ctttcaccac aactgctgag cgtgagattg ttagagacat gaaggagaag     720
54 ctctcttaca ttgccttgga ctttgaacaa gagctcgaga cttccaaaac aagctcatcc     780
56 gttgagaaga gcttcgagct gccagacggt caagtgatca ccatcggggc agagcgtttc     840
58 cgatgccctg aagttctgtt tcagccatcg atgatcggaa tggaaaatcc gggaattcat     900
60 gaaactactt acaactcaat catgaaatgt gatgtggata tcaggaagga tctttatgga     960
62 aacattgtgc ttagtggtgg caccacaatg ttcgatggga ttggtgatag gatgagtaaa    1020
64 gagatcacag cgttggctcg aagcagtatg aacatcaaag tggtggctcc accggaaagg    1080
66 aagtacagtg tctggatcgg tggctctatc ttggcttccc tcagtacttt ccagcaggta    1140
68 aattacttac tatacttaat acataaagtc tattagtgat ttgatgtata aagtgttaca    1200
70 aaaatgtgtt ccaaatttgc agatgtggat tgcgaaagcg gagtatgatg aatctggacc    1260
72 gtcaatcgtc cacaggaagt gcttctgatc aaaagtcacc aagtaaaaca agagcggtaa    1320
74 aaattttgat atcagttttt caccctgaag ccagttgcta taattactca caacttctct    1380
76 atttgtgttc ttttattctt gtccctcgtt gttcatttta atctcttttt tgcaacaaag    1440
78 caacttaaaa aaacagagca gtcattaaca gaatgttatt attatatata tgtatacata    1500
80 ttagtataca cccattattt cattaaaaca tttatcatat aaggatagga ttctatacat    1560
82 cgatatattt attttgttga cactattcag cacatgctta tgtcttatct tgttagtata    1620
84 tgtaaccaaa gacaaataat agatgctaca aattgttttc tttgaagcaa aaatttcaat    1680
86 cttaaaattg tttttttcca ggttacacaa aaaaaacttg tagtttgtaa attttctata    1740
```

Input Set : **A:\SeqList.txt**
Output Set: **N:\CRF3\05212001\I698213A.raw**

```
 88 caattttggg gatctcaaca agaacatgaa cttcaacttc tagtcatatg acgacctgag   1800
 90 tctgcgcggc tgtgaatctc tttgctgcag taaatgttta caagtggtgt gtaaattggt   1860
 92 actgattcaa aagctttaag aaatctacac atttcgtgaa attatttagc agacttgata   1920
 94 ttaaaaatct aggataaaat gactatccaa agacaaatag gactgtttca catgttcccc   1980
 96 tgattcttgt agctcataac tcatcagcag ttaactttc tacctcatac acgctcgcaa   2040
W--> 98 tncgtttgga attatcagct ntaatttttc taattctttg gaaattatta gcagctcgat   2100
100 caaatggggc atggcttctt cttctatctg caactcatct aaactttcca tgaagaaaca   2160
                                                                         2165
102 aagct
```

```
105 <210> SEQ ID NO: 2
106 <211> LENGTH: 6
107 <212> TYPE: PRT
108 <213> ORGANISM: Unknown
110 <220> FEATURE:
111 <223> OTHER INFORMATION: Describes a predicted protein sequence
113 <400> SEQUENCE: 2
114 Arg Phe Phe Arg Ala Leu
115 1               5
118 <210> SEQ ID NO: 3
119 <211> LENGTH: 7
120 <212> TYPE: PRT
121 <213> ORGANISM: Unknown
123 <220> FEATURE:
124 <223> OTHER INFORMATION: Describes a predicted protein sequence
126 <400> SEQUENCE: 3
127 Ala Val Leu Ala Thr Pro Val
128 1               5
131 <210> SEQ ID NO: 4
132 <211> LENGTH: 21
133 <212> TYPE: PRT
134 <213> ORGANISM: Unknown
136 <220> FEATURE:
137 <223> OTHER INFORMATION: Describes a predicted protein sequence
139 <400> SEQUENCE: 4
140 Trp Leu Gly Trp Asp Lys Arg Met Leu Met Leu Glu Thr Arg Leu Asn
141 1               5                   10                  15
143 Gln Asn Val Val Ser
144             20
147 <210> SEQ ID NO: 5
148 <211> LENGTH: 35
149 <212> TYPE: PRT
150 <213> ORGANISM: Unknown
152 <220> FEATURE:
153 <223> OTHER INFORMATION: Describes a predicted protein sequence
155 <400> SEQUENCE: 5
156 Ser Thr Gln Leu Ser Met Glu Leu Leu Ile Ile Gly Met Thr Trp Arg
157 1               5                   10                  15
159 Arg Phe Gly Ile Thr Leu Ser Thr Met Ser Phe Val Leu Pro Leu Lys
160             20                  25                  30
162 Asn Ile Arg
```

RAW SEQUENCE LISTING   DATE: 05/21/2001
PATENT APPLICATION: US/09/698,213A   TIME: 10:43:56

Input Set : A:\SeqList.txt
Output Set: N:\CRF3\05212001\I698213A.raw

```
163           35
166 <210> SEQ ID NO: 6
167 <211> LENGTH: 56
168 <212> TYPE: PRT
169 <213> ORGANISM: Unknown
171 <220> FEATURE:
172 <223> OTHER INFORMATION: Describes a predicted protein sequence
174 <400> SEQUENCE: 6
175 Leu Thr Glu Ala Pro Leu Asn Pro Lys Ala Asn Arg Glu Lys Met Thr
176 1               5                   10                  15
178 Gln Ile Met Phe Glu Thr Phe Asn Thr Pro Ala Met Tyr Val Ala Ile
179             20                  25                  30
181 Gln Ala Val Leu Ser Leu Tyr Ala Ser Gly Arg Thr Thr Gly Gln Tyr
182             35                  40                  45
184 Ile Thr Thr Phe Phe Leu Tyr Arg
185     50                  55
188 <210> SEQ ID NO: 7
189 <211> LENGTH: 191
190 <212> TYPE: PRT
191 <213> ORGANISM: Unknown
193 <220> FEATURE:
194 <223> OTHER INFORMATION: Describes a predicted protein sequence
196 <400> SEQUENCE: 7
197 Ser Gly Asp Gly Val Ser His Thr Val Pro Ile Tyr Glu Gly Tyr Ala
198 1               5                   10                  15
200 Leu Pro His Ala Ile Leu Arg Leu Asp Leu Ala Gly Arg Asp Leu Thr
201             20                  25                  30
203 Asp His Leu Met Lys Ile Leu Thr Glu Arg Gly Tyr Ser Phe Thr Thr
204             35                  40                  45
206 Thr Ala Glu Arg Glu Ile Val Arg Asp Met Lys Glu Lys Leu Ser Tyr
207     50                  55                  60
209 Ile Ala Leu Asp Phe Glu Gln Glu Leu Glu Thr Ser Lys Thr Ser Ser
210 65                  70                  75                  80
212 Ser Val Glu Lys Ser Phe Glu Leu Pro Asp Gly Gln Val Ile Thr Ile
213             85                  90                  95
215 Gly Ala Glu Arg Phe Arg Cys Pro Glu Val Leu Phe Gln Pro Ser Met
216             100                 105                 110
218 Ile Gly Met Glu Asn Pro Gly Ile His Glu Thr Thr Tyr Asn Ser Ile
219             115                 120                 125
221 Met Lys Cys Asp Val Asp Ile Arg Lys Asp Leu Tyr Gly Asn Ile Val
222     130                 135                 140
224 Leu Ser Gly Gly Thr Thr Met Phe Asp Gly Ile Gly Asp Arg Met Ser
225 145                 150                 155                 160
227 Lys Glu Ile Thr Ala Leu Ala Pro Ser Ser Met Lys Ile Lys Val Val
228             165                 170                 175
230 Ala Pro Pro Glu Arg Lys Tyr Ser Val Trp Ile Gly Gly Ser Ile
231             180                 185                 190
234 <210> SEQ ID NO: 8
235 <211> LENGTH: 13
```

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/698,213A

DATE: 05/21/2001
TIME: 10:43:56

Input Set : A:\SeqList.txt
Output Set: N:\CRF3\05212001\I698213A.raw

```
236 <212> TYPE: PRT
237 <213> ORGANISM: Unknown
239 <220> FEATURE:
240 <223> OTHER INFORMATION: Describes a predicted protein sequence
242 <400> SEQUENCE: 8
243 Val Pro Asn Leu Gln Met Trp Ile Ala Lys Ala Glu Tyr
244 1               5                   10
247 <210> SEQ ID NO: 9
248 <211> LENGTH: 26
249 <212> TYPE: PRT
250 <213> ORGANISM: Unknown
252 <220> FEATURE:
253 <223> OTHER INFORMATION: Describes a predicted protein sequence
255 <400> SEQUENCE: 9
256 Asn Leu Asp Arg Gln Ser Ser Thr Gly Ser Ala Ser Asp Gln Lys Ser
257 1            5                   10                  15
259 Pro Ser Lys Thr Arg Ala Val Lys Ile Leu
260             20                  25
263 <210> SEQ ID NO: 10
264 <211> LENGTH: 56
265 <212> TYPE: PRT
266 <213> ORGANISM: Unknown
268 <220> FEATURE:
269 <223> OTHER INFORMATION: Describes a predicted protein sequence
271 <400> SEQUENCE: 10
273 Asn Ser Ser Ala Val Asn Phe Ser Thr Ser Tyr Thr Leu Ala Ile Arg
274 1               5                   10                  15
276 Leu Glu Leu Ser Ala Leu Ile Phe Leu Ile Ser Leu Glu Ile Ile Ser
277             20                  25                  30
279 Ser Ser Ile Lys Trp Gly Met Ala Ser Ser Ser Ile Cys Asn Ser Ser
280         35                  40                  45
282 Lys Leu Ser Met Lys Lys Gln Ser
283     50                  55
286 <210> SEQ ID NO: 11
287 <211> LENGTH: 194
288 <212> TYPE: PRT
289 <213> ORGANISM: Unknown
291 <220> FEATURE:
292 <223> OTHER INFORMATION: Describes a predicted protein sequence
294 <400> SEQUENCE: 11
295 Ser Gly Asp Gly Val Ser His Thr Val Pro Ile Tyr Glu Gly Tyr Ala
296 1               5                   10                  15
298 Leu Pro His Ala Ile Leu Arg Leu Asp Leu Ala Gly Arg Asp Leu Thr
299             20                  25                  30
301 Asp His Leu Met Lys Ile Leu Thr Glu Arg Gly Tyr Ser Phe Thr Thr
302         35                  40                  45
304 Thr Ala Glu Arg Glu Ile Val Arg Asp Met Lys Glu Lys Leu Ser Tyr
305     50                  55                  60
307 Ile Ala Leu Asp Phe Glu Gln Glu Leu Glu Thr Ser Lys Thr Ser Ser
```

```
308 65                     70                 75                 80
310 Ser Val Glu Lys Ser Phe Glu Leu Pro Asp Gly Gln Val Ile Thr Ile
311             85                 90                 95
313 Gly Ala Glu Arg Phe Arg Cys Pro Glu Val Leu Phe Gln Pro Ser Met
314             100                105                110
316 Ile Gly Met Glu Asn Pro Gly Ile His Glu Thr Thr Tyr Asn Ser Ile
317             115                120                125
319 Met Lys Cys Asp Val Asp Ile Arg Lys Asp Leu Tyr Gly Asn Ile Val
320      130                135                140
322 Leu Ser Gly Gly Thr Thr Met Phe Asp Gly Ile Gly Asp Arg Met Ser
323 145                150                155                160
325 Lys Glu Ile Thr Ala Leu Ala Pro Ser Ser Met Lys Ile Lys Val Val
326             165                170                175
328 Ala Pro Pro Glu Arg Lys Tyr Ser Val Trp Ile Gly Gly Ser Ile Leu
329             180                185                190
331 Ala Ser
334 <210> SEQ ID NO: 12
335 <211> LENGTH: 9
336 <212> TYPE: PRT
337 <213> ORGANISM: Unknown
339 <220> FEATURE:
340 <223> OTHER INFORMATION: Describes a predicted protein sequence
342 <400> SEQUENCE: 12
343 Gln Met Trp Ile Ala Lys Ala Glu Tyr
344 1               5
347 <210> SEQ ID NO: 13
348 <211> LENGTH: 296
349 <212> TYPE: PRT
350 <213> ORGANISM: Arabidopsis thaliana
352 <220> FEATURE:
353 <223> OTHER INFORMATION: Ecotype columbia, describes actin
355 <400> SEQUENCE: 13
357 Met Glu Lys Ile Trp His His Thr Phe Tyr Asn Glu Leu Arg Val Ala
358 1               5                 10                 15
360 Pro Glu Glu His Pro Val Leu Leu Thr Glu Ala Pro Leu Asn Pro Lys
361             20                 25                 30
363 Ala Asn Arg Glu Lys Met Thr Gln Ile Met Phe Glu Thr Phe Asn Thr
364             35                 40                 45
366 Pro Ala Met Tyr Val Ala Ile Gln Ala Val Leu Ser Leu Ala Ser Gly
367      50                 55                 60
369 Arg Thr Thr Gly Gly Ile Val Leu Asp Ser Gly Asp Gly Val Ser His
370 65                 70                 75                 80
372 Thr Val Pro Ile Tyr Glu Gly Tyr Ala Leu Pro His Ala Ile Leu Arg
373             85                 90                 95
375 Leu Asp Leu Ala Gly Arg Asp Leu Thr Asp His Leu Met Lys Ile Leu
376             100                105                110
378 Thr Glu Arg Gly Tyr Ser Phe Thr Thr Thr Ala Glu Arg Glu Ile Val
379             115                120                125
381 Arg Asp Met Lys Glu Lys Leu Ser Tyr Ile Ala Leu Asp Phe Glu Gln
```

VERIFICATION SUMMARY   DATE: 05/21/2001
PATENT APPLICATION: US/09/698,213A   TIME: 10:43:57

Input Set : A:\SeqList.txt
Output Set: N:\CRF3\05212001\I698213A.raw

L:98 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1